

## INSECT IMMUNITY – A GENE SYSTEM FOR ANTIBACTERIAL PROTEINS

H.G. BOMAN, I. FAYE, R. GAN, G.H. GUDMUNDSSON, D.-A. LIDHOLM, J.-Y. LEE & K.G. XANTHOPOULOS

Department of Microbiology, University of Stockholm, S-106 91 Stockholm, Sweden

All higher organisms can in some way defend or protect themselves against their natural flora of bacteria. This applies also to insects which have highly effective immune systems, both cellular and humoral. Many insects respond to an injection of live, non-pathogenic bacteria with the production of a potent cell-free antibacterial activity (for a review see Boman & Hultmark, 1987). This phenomenon can be analysed at the molecular level using as a model system the Cecropia moth, *Hyalophora cecropia*. When a diapausing pupa is immunized it turns on predominantly the genes for immunity while the rest of the animal remains in a dormant state. Immunized pupae of Cecropia are therefore a system for biological enrichment of the RNA and the proteins which are synthesized from the genes for immunity. We have taken advantage of this fact both in the purification of 15 different immune proteins and in the isolation of immune RNA, later to be used for the preparation of a cDNA bank. After a short period of RNA synthesis the insects respond to live bacteria by the production of a potent antibacterial activity which is due to the synthesis of 15-20 immune proteins. To this group of proteins belong a lysozyme and two novel classes of antibacterial proteins, the attacins and the cecropins. The attacins contain 188 amino acid residues ( $M_r$  around 20,000) and there are two main forms, basic and acidic or neutral. The antibacterial spectra of the attacins are narrow. The cecropins are smaller and the three principal cecropins, A, B, and D, contain 35-37 amino acid residues ( $M_r$  around 4,000). The A and B forms are potent broad spectrum antibacterial agents.

During the years 1980-1985 parallel sequence work was carried out on the protein and on the DNA levels. This programme has produced the complete amino acid sequences for five cecropins, one lysozyme and one attacin. We first obtained cDNA sequences corresponding to one lysozyme, one cecropin (the B from) and the two major forms of attacin. More recently we have isolated cDNA clones for the two other cecropins and genomic clones for cecropin B. The present paper is a summary of

this structural work supplemented with data concerning the antibacterial activities for natural and synthetic cecropins.

### LYSOZYME

The first antibacterial factor to be identified in insect haemolymph was lysozyme and it was also the first factor to be purified (Powning & Davidson, 1973). It has been claimed that lysozyme is the main antibacterial factor responsible for immunity of vaccinated insects but this is not correct because insects can eliminate many lysozyme resistant bacteria. The Cecropia lysozyme was isolated in connection with our purification of cecropins A and B (Hultmark et al., 1980) and some years later the full amino acid sequence was worked out for the protein (Engström et al., 1985). A cDNA clone was also isolated for lysozyme and Fig. 1 shows its nucleotide sequence together with the deduced amino acid sequence (Engström et al., 1985). The enzyme is composed of 120 amino acids, has a size of 13.8 kD and

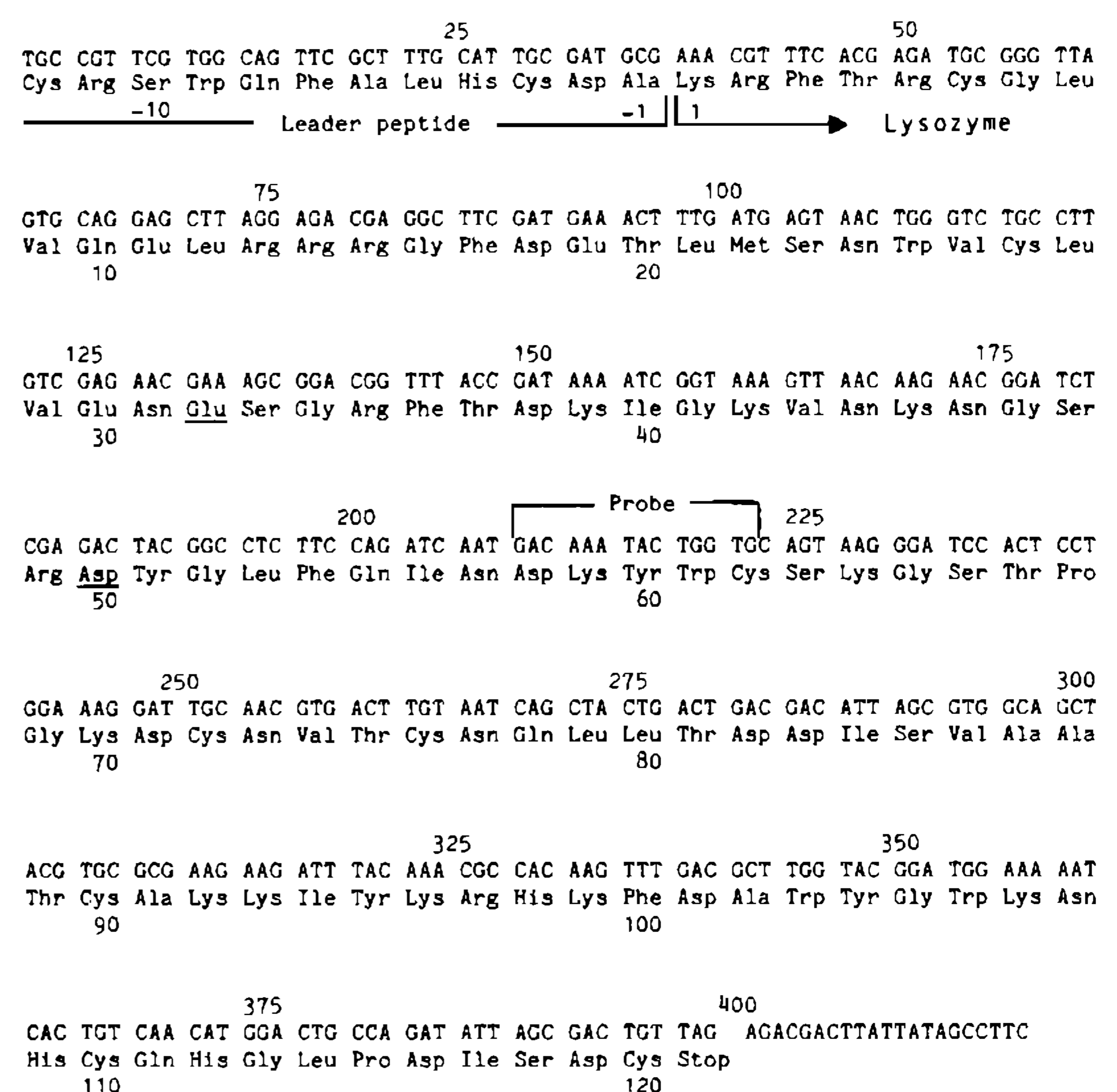


Fig. 1. Nucleotide and deduced amino acid sequences of the Cecropia lysozyme. Part of the leader peptide is indicated by a line under the respective sequence. The principal structure of the oligonucleotide probe (without the alternative wobble bases) is indicated by a line above the respective sequence. The active site of enzyme, residues Glu-32 and Asp-50, are underlined.

shows great similarity with vertebrate lysozymes of the chicken type. The amino acid residues responsible for the catalytic activity, for the S-S bridges and for the binding of substrate are essentially conserved. When the sequence translated from the cDNA clone was compared to the amino acid sequence of the protein there was an almost complete agreement. However, at two positions, 15 and 66, there were differences which most likely are due to allelic variations in the *Cecropia* population (Engström et al., 1985).

Lysozyme is bactericidal to only some Gram positive bacteria like *Bacillus megaterium*, *Bacillus subtilis* and *Micrococcus luteus*. However, cecropins A and B also act on two of these bacteria and *B. subtilis* is so far the only bacterium that is fairly sensitive to lysozyme and rather resistant to cecropins. Thus, the main function of the lysozyme may not be to kill sensitive bacteria but to remove the murein sacculus which is left after the action of cecropins and attacins.

#### TWO MAIN FORMS OF ATTACIN

The attacins were first isolated by molecular sieving as an antibacterial fraction with molecular weight considerably larger than the cecropins (Hultmark et al., 1983). Subsequent studies revealed as many as six different components (A-F) which could be fractionated according to

isoelectric point. To our surprise they all turned out to react with antisera prepared against our immune protein P5 isolated several years earlier (Pye & Boman, 1977). At that time we had not been able to find any antibacterial activity of P5, a discrepancy that was explained by the later finding that attacins only act on growing bacteria (Hultmark et al., 1983).

An Ouchterlony immuno-diffusion plate revealed first that all six attacins shared one antigenic determinant, while another one was common only to attacins E and F, the two acidic forms. The N-terminal sequences for five of the attacins indicated that the three basic forms all have similar sequences while the two acidic forms are identical but slightly different from the basic (Hultmark et al., 1983). These data strongly suggested the existence of only two different genes, one for the basic and one for the neutral or acidic form. This was born out by the isolation and sequencing of two cDNA clones, pCP517 and pCP521 (Kockun et al., 1984). Fig. 2 shows that the two main attacins are very similar with as much as 79% homology at the amino acid level. On the DNA level the homology is 76% for the coding region in contrast to only 36% in the region beyond the stop signal.

Engström et al. (1984a) worked out the full amino acid sequence for attacin F. There is a complete agreement between these protein data

TABLE  
Lethal Concentrations ( $\mu\text{M}$ ) of Cecropins on four test bacteria

Compound Origin	Bacterium and strain			
	<i>E. coli</i> D21	<i>P. aeruginosa</i> OT97	<i>B. megaterium</i> Bm11	<i>M. luteus</i> M111
Cecropin A natural	0.34	3.1	0.60	1.6
Cecropin A synthetic	0.36	2.4	0.81	1.8
Cecropin B natural	0.52	1.6	0.95	1.5
Cecropin B synthetic	0.66	1.4	0.75	1.3
Cecropin D natural	3.6	>108	26	37
Cecropin D synthetic	3.1	66	31	12

Thin agar plates were seeded with the respective test bacteria. Small wells were punched in the plates and loaded with a dilution series of each sample. After overnight incubation at 30°C the inhibition zones were recorded and lethal concentrations calculated as described by Hultmark et al (1983). Since large zones give low values for lethal concentration, low figures in the table are in general more accurate than high values. The experimental errors in the method come chiefly from the determination of the concentration of the cecropin, from pipetting errors and from the quality of the plastic plate. Using high quality plates and the same series of diluted samples the error is often less than 20% for a high activity sample. In case of cecropin D acting on *M. luteus* the zones were very diffuse and difficult to record. Here we believe the error may be a factor of 2-3.

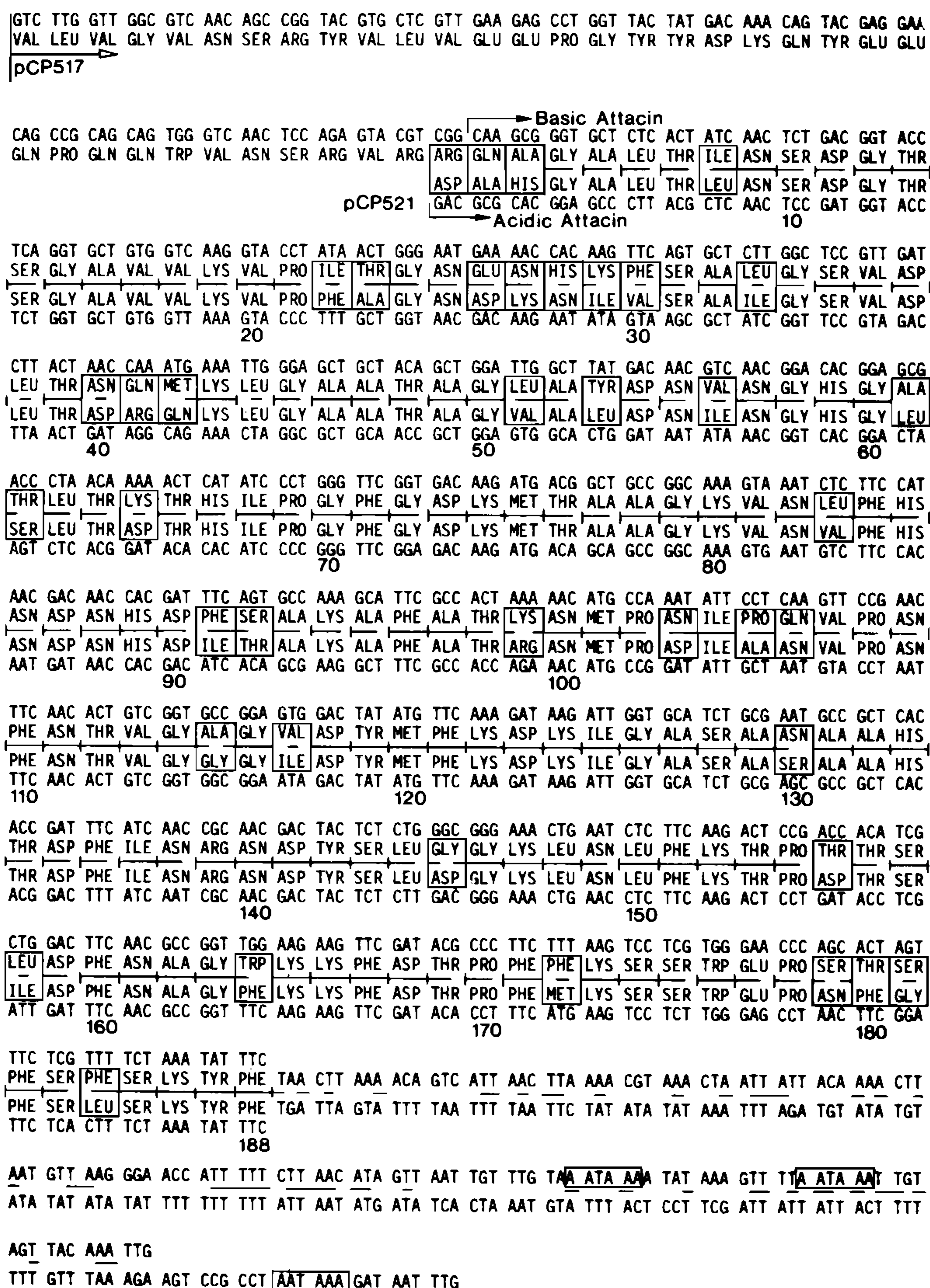


Fig. 2. Nucleotide sequences of the inserts in two attacin clones, pCP517 (upper-most line) and pCP521 (bottom line). The amino acid sequence corresponding to pCP517 is given below the nucleotide sequence, for pCP521 above. The line between the amino acid sequences is broken when the nucleotide sequences differ. Amino acid substitutions and the signals for polyadenylation are boxed in. The numbers indicate amino acid positions for both attacins.

and amino acid residues 1-184 deduced from pCP521. However, pCP521 codes for an extended protein of 188 amino acid residues (attacin E). The difference corresponds to a C-terminal tetrapeptide, Ser-Lys-Tyr-Phe, which is also coded for in the clone for the basic attacin. Since this peptide contains one positive charge, an incomplete proteolytic removal could account for four attacins with different charges. It cannot be decided at present if such a processing is the result of an artificial proteolytic cleavage or if it has a natural function.

In the case of pCP517 we obtained 36 amino acid residues of the leader sequence.

Since neither the length nor the composition of this region is typical for a signal sequence, we believe attacins to be made from a pre-pro-form. Similar conclusions had earlier been reached from the size of the proteins produced by *in vitro* translation of immune mRNA (Lee et al., 1983). In addition there may be a further trimming also at the N-terminus of the basic attacin because evidence for a pyroglutamate group was found at the N-terminus of the basic attacin (Engström et al., 1984a). We therefore believe that the mature protein starts with the Gln residue indicated in Fig. 2. A partial cyclization of this residues would involve a partial

loss of one positive charge. An enzyme that can perform this reaction was recently identified (Fisher & Spiess, 1987). Thus, together with the removal of the tetrapeptide at the C-terminus it is possible by different steps of processing to account for all six forms of attacins observed.

The antibacterial spectra of the attacins were found to be narrow with good activity only against *E. coli* and two other bacteria originating from the gut of an *Antheraea* larva (Hultmark et al., 1983). A study of the mechanism of action on *E. coli* demonstrated that the two main attacins both act on the outer membrane (Engström et al., 1984b). In particular it was shown that attacin facilitates the action of cecropin and lysozyme, thereby enabling these three immune proteins to work in consonance.

### THREE MAIN CECROPINS AND THEIR PRECURSOR FORMS

Cecropins were discovered in 1979 when we succeeded in their separation from the *Cecropia* lysozyme. Once separated, cecropins A and B were isolated simultaneously with the lysozyme (Hultmark et al., 1980). Two years later we found cecropin D as well as some minor forms believed to be precursors with an additional Gly residue (Hultmark et al., 1982). In collaboration with the Shanghai Institute of Biochemistry, we have also isolated cecropins D and B from the Chinese oak silk moth, *Antheraea pernyi* (Qu et al., 1982). The amino acid sequences for these five small proteins showed that they all were in principle similar type of molecules with a strongly basic N-terminal region and a long hydrophobic stretch in the C-terminal half (for details see Boman & Hultmark, 1987).

It turned out to be difficult to find cecropin clones using synthetic oligonucleotide probes designed from the known amino acid sequences. Our first search gave only three similar clones of which the two larger were identified as coding for cecropin B (v Hofsten et al., 1985). The overlapping sequences showed an open reading frame with 62 amino acid residues of which the mature cecropin B only represented 35. After almost two years of additional work (and some frustration) clones were obtained also for cecropins A and D. The sequences of the coding regions of these clones are presented in Figs. 3 and 4 (data from Lidholm et al., 1987). In the former we compare the prepro-forms of cecropins A and B, in the latter the correspond-

ing A and the D forms. These data brought us several new pieces of information.

Firstly, all three cecropins are made as larger precursors with sequences of 24-26 amino acid residues in the N-terminus which are not present in the mature proteins. The signal peptides probably makes up the first 22 of these 24-26 amino acids, leaving one or two Pro-containing dipeptides before the start of the mature cecropin B. Of special interest is the fact that the first 12 residues of the signal peptide is fully conserved in the precursors of cecropins A and B. For the N-terminus of the D-form 8 of 13 residues are shared with the A-form and the remaining 5 residues are all conservative replacements. This region was conserved also in the meat fly *Sarcophaga* (Matsumoto et al., 1986). However, the regions between residues -14 and -8 were not conserved in any of the cecropins (see Figs. 3 and 4). Together these findings are unusual and the conserved N-terminal region must have a function, possibly as a recognition part for a receptor mediated transport, specially designed for very small proteins.

Secondly, the tetrapeptide, Ala-Pro-Glu-Pro, present in the proforms of cecropins A and B is identical to the first four amino acid residues of the prosequence of melittin (Kreil et al., 1980). Oddly enough, there is in the D-form a three-residue deletion that has removed the second Pro-containing dipeptide and the first residue of the mature cecropin. (Alternatively an insertion in the ancestor of A and B forms is possible.) Since no signal peptidase was found to cleave a Pro-Lys bond (v Heine, 1985), it seems likely that cecropins, like melittin, are processed by a dipeptidyl peptidase (Kreil et al., 1980). In concistence is the finding that both hemolymph and fatbody of *Cecropia* contain enzyme(s) that split Ala-Pro-nitroanelide (Boman unpublished).

Thirdly, the amide group in the C-terminus of the mature cecropins must be derived from the Gly residue which in the A and B forms terminates the coding part of the cDNA-sequence (Figs. 3 and 4). The mechanism of amidation is thus analogous to the one suggested for melittin (Vlasak et al., 1983) and several brain hormones (Lynch & Snyder, 1986). The corresponding in vitro reaction has also been demonstrated (Bradbury et al., 1982; Elipper et al., 1983). Mature cecropin D ends with a blocked Lys residue while the clone codes for two additional amino acids, a Gly and a Lys. If the amidating enzyme works on Gly-Lys is not

**COMPARISON OF PREPRO-FORMS OF CECROPINS A AND B**

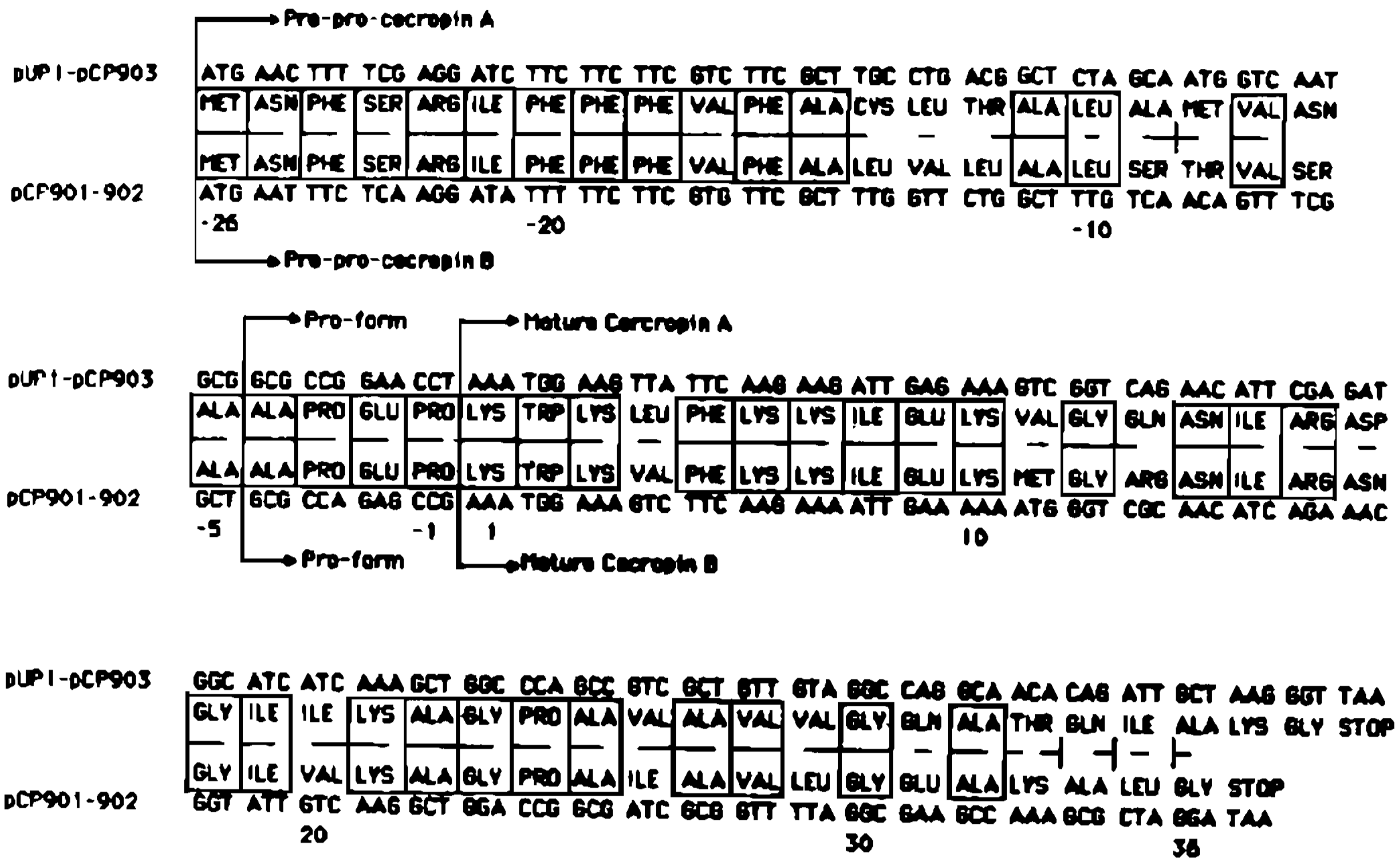


Fig. 3. Nucleotide sequences of cDNA clones for cecropins A and B. The translation of the upper line of nucleotides is given on the line below while the reverse is the case for the lower line of nucleotides. Identical amino acid residues are boxed in and the line between the amino acid residues indicates identical nucleotides, when broken a case of nucleotide substitution.

**COMPARISON OF PREPRO - FORMS OF CECROPINS A and D**

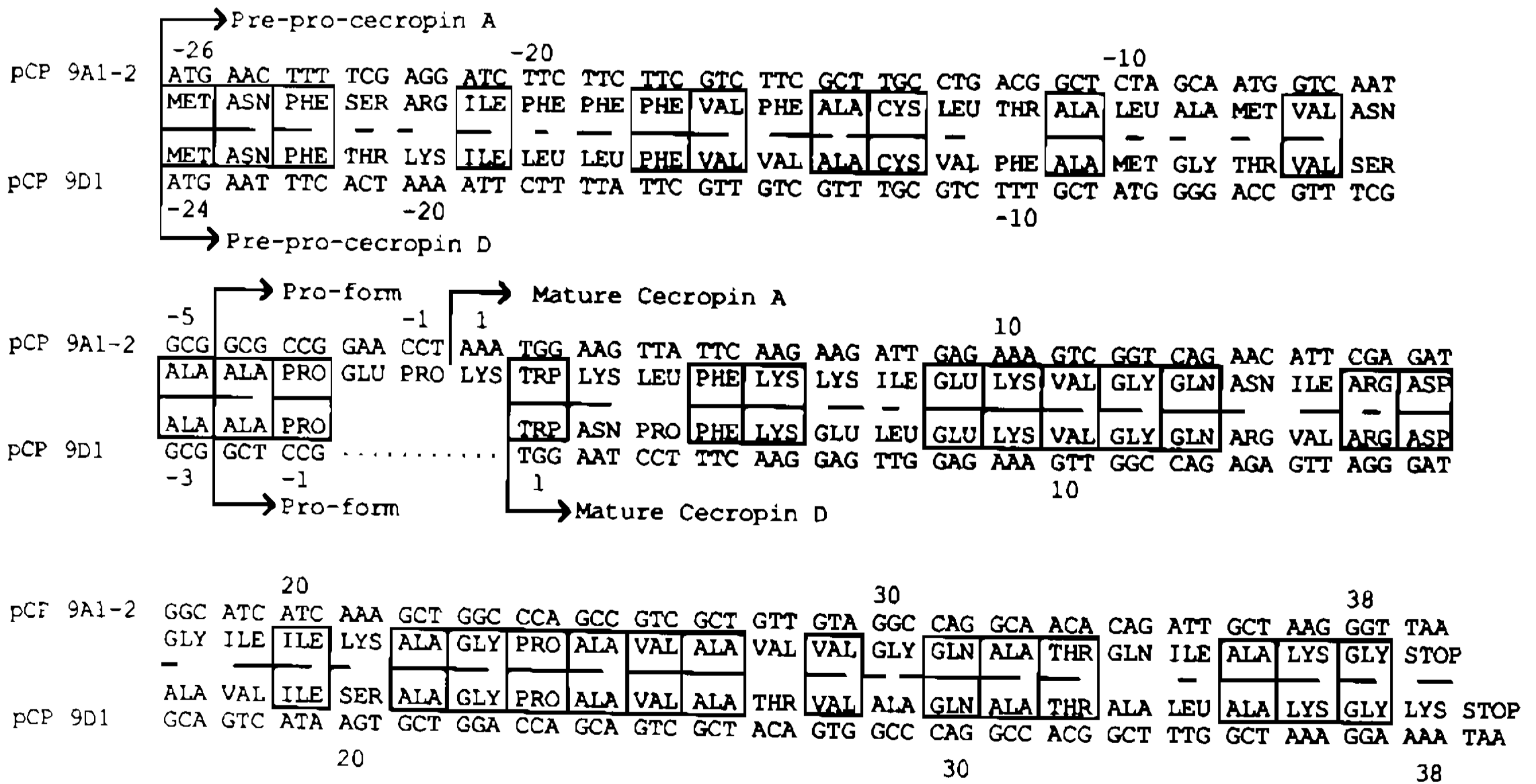


Fig. 4. Nucleotide sequences of cDNA clones for cecropins A and D. The translation of the upper line of nucleotides is given on the line below while the reverse is the case for the lower line of nucleotides. Identical amino acid residues are boxed in and the line between the amino acid residues indicates identical nucleotides, when broken a case of nucleotide substitution.

known. If not, we will have to involve in the processing also a carboxypeptidase H (Lynch & Snyder, 1986) for the removal of the terminal Lys residue. In summary we can conclude that in order to obtain the mature cecropins the respective precursor molecules will have to be processed in three to four steps at both ends.

#### SOLID-PHASE SYNTHESIS OF CECROPINS

In collaboration with Bruce Merrifield's laboratory at the Rockefeller University a synthetic programme was initiated in order to confirm the structure of the cecropins. The program was then continued with the aim of investigating a possible correlation between the secondary structure and the antibacterial activity. A beginning was made with the synthesis of cecropin A(1-33), at that time thought to be the full cecropin A (Merrifield et al., 1982). This work was followed by the synthesis of the complete cecropin A and some truncated analogues (Andreu et al., 1983). Also cecropin B was synthesized with a revised C-terminus deduced from our cDNA work (v Hofsten et al., 1985). For both cecropins the C-terminal peptide was isolated and the amidation was confirmed by mass spectroscopy. Most recently, the D-form was synthesized by Fink and Merrifield and found to have the same antibacterial spectra as the natural compound. These data are given in the Table which also shows the broad spectrum properties of cecropins A and B.

In the N-terminal region of the mature cecropins, polar and hydrophobic side chains are interspaced in a regular pattern. Steiner (1982) and Merrifield et al. (1982) observed that this amino acid distribution is likely to produce an amphipathic alpha helix. Such structures have often been implicated in membrane activities and the cecropins are indeed strongly lytic against a variety of bacteria. Cecropins will also lyse artificial liposomes (Steiner et al., *Biochim. Biophys Acta*, in press 1988).

As a continuation of the synthetic programme a series of analogues of cecropin A were synthesized in which residues 2, 6 and 8 were altered in such a way that the nature of the side chain was changed from hydrophobic to hydrophilic or vice versa (Andreu et al., 1985). We also replaced residues 4 or 8 by Pro in order to break the alpha helix. From this work we could conclude that Trp-2 is an essential residue and that the antibacterial mechanism as such does not require an extended alpha helix. However,

the two Pro analogs did not show any activity against *M. luteus* so the broad spectrum activity was lost when the alpha helix was disrupted.

#### SEQUENCE OF A GENE FOR CECROPIN B

About two years ago we prepared a genomic bank from 10 *Cecropia* pupae using Charon 4A as vector. With one of our cDNA clones for cecropin B as probe we isolated and mapped four genomic clones for the B form (Xanthopoulos et al., *Eur. J. Biochem.*, in press 1988). Restriction mapping showed these clones to be similar only in the proximity of the coding region, otherwise they did not resemble each other. The reason for this is at present unclear: we could have multiple copies of the gene present in one animal and we could have multiple alleles in the *Cecropia* population.

The transcriptional region of one genomic cecropin B clone was sequenced and the result is given in Fig. 5. There is a sequence homologous to the TATA box at position -30 and a CAT box like sequence at position -68 (Breathnach & Chambon, 1981). Transcription for the cecropin B gene starts at position +1 with a heptanucleotide ATCATTTC. Many insect mRNAs have a  $\text{ATCAG}_{\text{T}}\text{TC}_{\text{T}}$  consensus sequence at the 5' end of the genes (Snyder et al., 1982; Hultmark et al., 1986). Finally, in the 3' end of the gene there are two AATAAA sequences homologous to the polyadenylation signal (Proudfoot & Brownlee, 1976).

The coding sequence of cecropin B gene is interrupted by a single intron located between amino acid residues eight and nine of the mature molecule and this area is highly conserved in the eight cecropin sequences so far available (Boman & Hultmark, 1987).

An overview of the gene for cecropin B and the primary translation product is given in Fig. 6. Preprocecropin B with its 62 amino acid residues is rather short to be a primary translation product. In case of melittin, a molecule of 26 amino acid residues, the pre-pro form is 70 amino acid residues long (Suchanek et al., 1978). Similarly, the common precursor of human pancreatic polypeptide and pancreatic isosapeptide is 95 amino acids long, while the mature products are only 36 and 20 amino acids long, respectively (Boel et al., 1984). Also the antifreeze protein component A of the winter flounder, a 37 amino acid residues peptide, is the final product of a 82 residues long precursor (Davies et al., 1984). Many other low molecular weight proteins are synthesized as

-562	AGATCTATTG	GAACTGATAT	AAACACCATT	AATACACATT	AAGATATTAA
-512	ATTTTATGGT	TTTGTTCCGA	ACAGAATTAA	CAATTTAAAT	TCGAATAAAA
-462	AATATTACCG	TTGTAAAATA	CTCGTGATTT	TACGACATGT	GTTAATATTA
-412	AAATTGGAAG	CTCTCGTTTA	AACAATTTTG	AAATATTTT	CTGTGGTGTC
-362	CAAGGAGGGT	TTAATAAAAA	AGTTCCAAAC	TCTTTGAAAA	AGTAACTACC
-312	ATAATAATGT	AGTAGATACA	TATGTAGTAA	ATATTATACG	TATAAATGTG
-262	CCGTTGGACA	TAATTTACAA	TTAAAAAAA	TTACGTCGAT	TTCATTTGAA
-212	TTTTGACCCC	GGTCAGCTTC	CTTTTACGTT	GCGAGGGAAT	TTACATTTAT
-162	CTACGCTGTA	ACATCCAAAC	TGTTTACATT	TGACCCCAA	TAAGCGGTTA
-112	TCAGACTTGA	CTCCGCTGCA	TAAGTGCTAA	TTTAATAAAC	TCTTGCAAAT
-62	GTCTGCGTCA	TTTATTGTAG	CAAAGATTTT	<u>AATATAAAC</u>	AGACAATTAA
-12	TTTATTTCGAT	AC <span style="border: 1px solid black; padding: 2px;">ATCATTC</span> A	TCTCGTGACT	TCTTCGTGTG	TGGTGTTTAC
		+1 cap site			
39	CTATATATCT	AAATTTAATA	TTTCGTTTAT	TAAAATTTAA	TATATTTCGA
89	CGATGAATTT MetAsnPh	CTCAAGGATA eSerArgIle	TTTTTCTTCG PhePhePheV	TGTTCGCTTT aPheAlaLe	GGTTCGGCT uValLeuAla
			<b>Exon I</b>		
139	TTGTCAACAG LeuSerThrV	TTTCGGCTGC aSerAlaAl	GCCAGAGCCG aProGluPro	AAATGGAAAG LysTrpLysV	TCTTCAAGAA aPheLysLy
189	AATTGTAAGT sIle	TTATTTTATA	TTTACTTAA	TAGTGTTTAG	ATCTGTATTT
239	AATGATCCAT	TTTAAAAAGT	ACACTTTTGA	AATTAAGACG	ACGGATTTT
289	ATTTCTATTA	GCCACAAATA	CAATCGAATT	TCTCAATCGA	AAGTATTTT
339	CGACATCTCT	GAGGTAGCTT	TAGCTAGAAA	AAATATTTAT	TTTTAAAAA
389	TATTTTTTTG	TTGCGTTTCT	GATTTAAGTA	TGTTTATTTA	CAAAGTATG
439	TATGCCAAAT	TATCCTCAAT	TTATATATCA	GCCAAGTTTT	CAAAGTCTT
489	AGAAAGTAAA	TATAATATTA	TTCTTTAAAG	CTATCAACAG	CTATATTTGT
539	ACTAACTGTT	CTTGGATATT	TTTTCGTTTG	CAAATACTC	TATTTTTTAA
589	ATTTGTGTTA	ATATCATATT	GTGGTACTT	TTTTTCATAT	TCAAATAAC
639	TTCAAAGAAA	GGCTAATAAT	GGTAACTTT	TGAAACCGTA	CCTAATAAAT
689	TTGTGTTTTT	TTTTCCAGGA Gl	AAAATGGGT uLysMetGly	CGCAACATCA ArgAsnIleA	GAAACGGTAT rgAsnGlyIl
			<b>Exon II</b>		
739	TGTCAAGGCT eValLysAla	GGACCGGCGA GlyProAlaI	TCGCGGTTTT leAlaValLe	AGGCGAAGCC uGlyGluAla	AAAGCGCTAG LysAlaLeuG
789	GATAAAATAA ly***	TTTTAATTTA	AAATATTATT	TATTGATAAA	CGTTTTTGTT
839	ACTATTATAT	TATTTAATTT	AGATAATAAA	TTTAATTTAT	AAATTTTCAT
889	TGTTAATAAT	TTAATTTGTC	CTTTAATAAT	AGGTTTAATA	ACAGGACATC
939	CTTTATACCT	TGCGTGCGTT	<u>TGAAAATAAA</u>	CTTTATTTAA	TGT

Fig. 5. Nucleotide sequence of the cecropin B gene of clone  $\lambda$ CP9B12. Sequences underlined are homologous to the TATA box, the CAT box and the two polyadenylation signals AATAAA. The cap site is boxed. Triangles indicate the presumptive sites of cleavage of the precursor molecule.

## CECROPIN B GENE

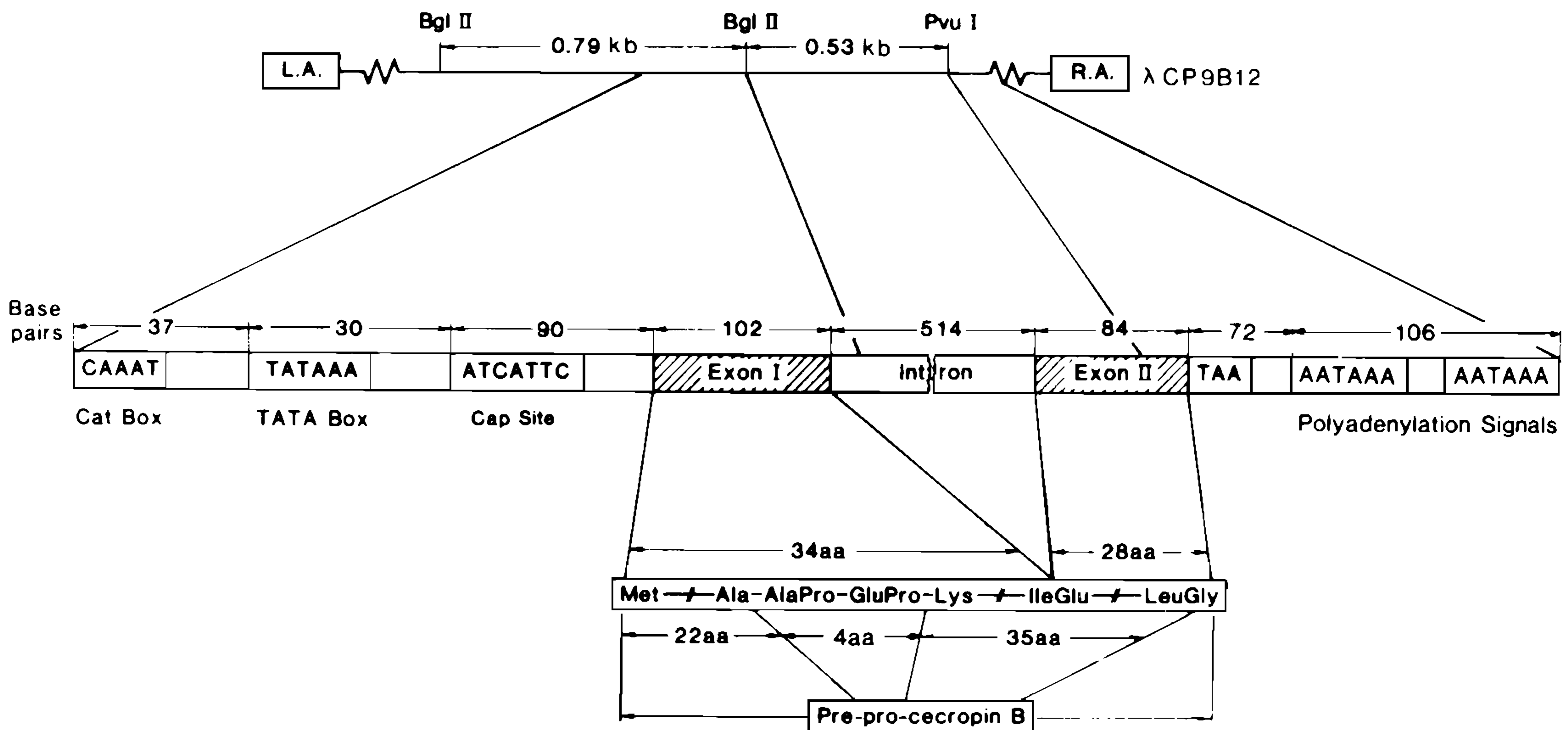


Fig. 6. An overview of the gene for cecropin B and its primary translation product. The different segments of the transcriptional unit are not drawn to scale. For preprocecropin B is given only the amino acid residues around cleavage or splice sites. The pro-form is believed to be processed by a dipeptidyl peptidase. The C-terminal glycine in preprocecropin B is split to give and amidated C-terminal leucine in mature cecropin B.

polyproteins which is the case also of the magainins, the 23 amino acid residues antibacterial peptides made in frog skin (Zasloff, 1987). This limited comparison indicates so far that the cecropin B precursor may be the smallest precursor molecule found and that small precursors cannot be smaller than 60-80 residues. A possible explanation is that the synthetic machinery creates a lower size limit for a primary translation product of an exported protein.

## DISCUSSION

The cecropins and the attacins were first defined by their separation properties, and in particular by the acidic electrophoresis in combination with an antibacterial assay on top of the gel (Hultmark et al., 1980). With this technique we early demonstrated cecropin-like substances in seven other Lepidopteran species. However, it was emphasized that ultimately it will be necessary to establish the identity of a cecropin by sequence analysis. A review of all known cecropins and all reports of cecropin-like compounds was recently given (Boman & Hultmark, 1987) and will not be repeated here.

So far, we have been the only group working on *Cecropia* immunity. It has therefore been an

important part of our strategy to confirm in an independent way as many structures as possible. For cecropin A the protein sequence was confirmed by solid-phase synthesis and fingerprinting of the natural and the synthetic material (Andrew et al., 1983). For all cecropins the cDNA sequences provide independent confirmation of the protein structures. This technique made it possible to detect an early error in the C-terminal sequence for cecropin B (see Steiner et al., 1981; v Hofsten et al., 1985).

The *Cecropia* moth has three different cecropins and at least two different attacins. The attacins are very similar with 79% homology on amino acid level. The mature cecropins differ somewhat more but the D and the B forms still show 62 and 65% homology with the A form. These data suggest that both cecropins and attacins have evolved through a series of gene duplications. So far we have not been able to document any real differences in function between these multiple antibacterial factors. In general, cecropin B is slightly more potent than the A-form, while cecropin D has a rather narrow antibacterial spectrum. However, in no case do we have a bacterium on which only a single factor acts. This raises the ques-



tion of the survival value of multiple forms of very similar molecules. One alternative is that they simply represent proteins "in the middle" of an evolution towards separate functions. A second possibility is that cecropins and attacins both have separate target organisms which we have just not found. A third alternative is that each of the cecropins and attacins have separate targets on most of the organisms on which they act. If so, this would in itself provide a survival value to the insect because it would make it virtually impossible for a susceptible bacterium to produce mutants which are resistant to the humoral immunity of an insect.

The specificity strongly suggest that in nature bacteria are the prime targets of the cecropins, attacins and lysozyme. At the same time the high growth rates of bacteria makes them evolve, perhaps 50-100 times faster than insects. It may be argued that in order to survive a rapidly changing selection pressure from different bacteria, the insects have developed the most adaptable antibacterial immune system that was possible. This may be the reason why we deal with a system of broad spectrum antibacterial molecules originating from gene duplications with multiple copies of each gene.

#### ACKNOWLEDGEMENTS

The synthetic work has been possible only through collaboration with Bruce Merrifield and his colleagues in New York. The work in Stockholm was supported by grants from the Swedish Natural Science Research Council.

#### REFERENCES

- ANDREU, D.; MERRIFIELD, R.B.; STEINER, H. & BOMAN, H.G., 1983. Solid-phase synthesis of cecropin A and related peptides. *Proc. Natl. Acad. Sci. U.S.A.*, *80* :6475-6479.
- ANDREU, D.; MERRIFIELD, R.B.; STEINER, H. & BOMAN, H.G., 1985. N-Terminal analogs of cecropin A: Synthesis, antibacterial activity and conformational properties. *Biochem.*, *24* :1683-1688.
- BOEL, E.; SCHWARTZ, T.W.; NORRIS, K.E. & FILL, N.P., 1984. A cDNA encoding a small common precursor for human pancreatic polypeptide and pancreatic icosapeptide. *EMBO J.*, *3* :909-912.
- BOMAN, H.G. & HULTMARK, D., 1987. Cell-free immunity in insects. *Ann. Rev. Microbiol.*, *41* :103-126.
- BRADBURY, A.F.; FINNIE, M.D.A. & SMYTH, D. G., 1982. Mechanism of C-terminal amide formation by pituitary enzymes. *Nature*, *296* :686-688.
- BREATHNACH, R. & CHAMBON, P., 1981. Organization and expression of eucaryotic split genes coding for proteins. *Ann. Rev. Biochem.*, *50* :349-383.
- DAVIES, P.L.; HOUGH, C.; SCOTT, G.K.; NG, N., WHITE, N. & HEW, C.L., 1984. Antifreeze protein genes of the winter flounder. *J. Biol. Chem.*, *259* :9241-9247.
- EIPPER, B.A.; MAINS, R.E. & GLEMBOTSKI, C.C., 1983. Identification in pituitary tissue of a peptide alfa-amidation activity that acts on glycine extended peptides and requires molecular oxygen, copper, and ascorbic acid. *Proc. Nat. Acad. Sci., USA*, *80* :5144-5148.
- ENGSTROM, Å., ENGSTRÖM, P.; TAO Z-J., CARLSSON, A. & BENNICH, H., 1984a. Insect Immunity. The primary structure of the antibacterial protein attacin F and its relation to two native attacins from *Hyalophora cecropia*. *EMBO J.*, *3* :2065-2070.
- ENGSTRÖM, P., CARLSSON, A., ENGSTRÖM, Å., Z.-j. & BENNICH, H., 1984b. The antibacterial effect of attacins from the silk moth *Hyalophora cecropia* is directed against the outer membrane of *Escherichia coli*. *EMBO J.*, *3* :3347-3351.
- ENGSTRÖM, Å., XANTHOPOULOS, K.G., BOMAN, H.G. & BENNICH, H., 1985. Amino acid and cDNA sequences of lysozyme from *Hyalophora cecropia*. *EMBO J.*, *4* :2119-2122.
- FISHER, W.H. & SPIESS, J., 1987. Identification of a mammalian glutaminyl cyclase converting glutaminyl into pyroglutamyl peptides. *Proc. Nat. Acad. Sci. U.S.A.*, *84* :3628-3632.
- v. HEIJNE, G., 1985. Signal sequences. The limits of variation. *J. Mol. Biol.*, *184* :99-105.
- v. HOFSTEN, P.; FAYE, I.; KOCKUM, K.; LEE, J.-Y.; XANTHOPOULOS, K.G.; BOMAN, H.G.; BOMAN, H.G.; ENGSTRÖM, Å., ANDREU, D. & MERRIFIELD, R.B., 1985. Molecular cloning, cDNA sequencing and chemical synthesis of cecropin B from *Hyalophora cecropia*. *Proc. Natl. Acad. Sci. U.S.A.*, *82* :2240-2243.
- HULTMARK, D.; STEINER, H.; RASMUSON, T. & BOMAN, H.G., 1980. Insect Immunity. Purification and properties of three inducible bactericidal proteins from hemolymph of immunized pupae of *Hyalophora cecropia*. *Eur. J. Biochem.*, *106* :7-16.
- HULTMARK, D.; ENGSTRÖM, Å., BENNICH, H.; KAPUR, R. & BOMAN, H.G., 1982. Insect Immunity. Isolation and structure of cecropin D and four minor antibacterial components from *Cecropia* pupae. *Eur. J. Biochem.*, *127* :207-217.
- HULTMARK, D.; ENGSTRÖM, Å., ANDERSSON, K.; STEINER, H.; BENNICH, H. & BOMAN, H. G., 1983. Insect Immunity. Attacins, a family of antibacterial proteins from *Hyalophora cecropia*. *EMBO J.*, *2* :571-576.
- HULTMARK, D.; KLEMENZ, R. & GEHRING, W. J., 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. *Cell* *44* :429-438.
- KOCKUM, K.; FAYE, I.; v. HOFSTEN, P.; LEE, J.-Y.; XANTHOPOULOS, K.G. & BOMAN, H.G., 1984. Insect immunity. Isolation and sequence of two cDNA clones corresponding to acidic and basic attacins from *Hyalophora cecropia*. *EMBO J.*, *3* :2071-2075.
- KREIL, G.; HAIML, L. & SUCHANEK, G., 1980. Stepwise cleavage of the propeptide of promelittin by Dipeptidylpeptidase IV. Evidence for a new type of precursor - product conversion. *Eur. J. Biochem.*, *111* :49-58.
- LEE, J.-Y.; EDLUND, T.; NY, T.; FAYE, I. & BOMAN, H.G., 1983. Insect Immunity. Isolation of

- cDNA clones corresponding to attacins and immune protein P4 from *Hyalophora cecropia*. *EMBO J.*, 2 :577-581.
- LIDHOLM, D.-A.; GUDMUNDSSON, G.H.; XANTHOPOULOS, K. & BOMAN, H.G., 1987. Insect Immunity. cDNA clones coding for the precursor forms of cecropins A and D, antibacterial proteins from *Hyalophora cecropia*. *FEBS letters*, 226 :8-12.
- LYNCH, D. R. & SNYDER, S.H., 1986. Neuropeptides: Multiple molecular forms, metabolic pathways and receptors. *Ann. Rev. Biochem.* 55 :773-799.
- MATSUMOTO, N.; OKADA, M.; TAKAHASHI, H.; MING, Q. X.; NAKAJIMA, Y.; NAKANISHI, Y., KOMANO, H. & NATORI, S., 1986. Molecular cloning of a cDNA and assignment of the C-terminal of sarcotoxin IA, a potent antibacterial protein from *Sarcophaga peregrina*. *Biochem. J.*, 239 :717-722.
- MERRIFIELD, R.B.; VIZIOLI, L.D. & BOMAN, H. G., 1982. Synthesis of the antibacterial peptide cecropin A (1-33). *Biochem.*, 21 :5020-5031.
- POWNING, R.F. & DAVIDSON, W.J., 1973. Studies on insect bacteriolytic enzymes. I. Lysozyme in haemolymph of *Galleria mellonella* and *Bombyx mori*. *Comp. Biochem. Physiol.*, 45B :669-681.
- PROUDFOOT, N. J. & BROWNLEE, G. G., 1976. 3' non-coding region sequences in eucaryotic mRNA. *Nature*, 263 :211-214.
- PYE, A.E. & BOMAN, H.G., 1977. Insect Immunity. III. Purification and partial characterization of immune protein P5 from haemolymph of *Hyalophora cecropia* pupae. *Infect. Immun.*, 17 :408-414.
- QU, X.-M.; STEINER, H.; ENGSTRÖM, Å.; BENNICH, H. & BOMAN, H.G., 1982. Insect immunity: isolation and structure of cecropins B and D from pupae of the Chinese oak silk moth, *Antheraea pernyi*. *Eur. J. Biochem.* 127 :219-224.
- SNYDER, M.; HUNKAPILLER, M.; YUEN, D.; SILVERT, D.; FRISTROM, J. & DAVIDSON, N., 1982. Cuticle protein genes of *Drosophila*: Structure organization and evolution of four clustered genes. *Cell*, 29 :1027-1040.
- STEINER, H.; HULTMARK, D.; ENGSTRÖM, Å.; BENNICH, H. & BOMAN, H.G., 1981. Sequence and specificity of two antibacterial proteins involved in insect immunity. *Nature.*, 292 :246-248.
- STEINER, H., 1982. Secondary structure of the cecropins: Antibacterial peptides from the moth *Hyalophora cecropia*. *FEBS Letters* 137 :283-287.
- SUCHANEK, G.; KREIL, G. & HERMODSON, M.A., 1978. Amino acid sequence of honeybee prepromelittin synthesized *in vitro*. *Proc. Natl. Acad. Sci. USA.*, 75 :701-704.
- VLASAK, R.; UNGER-ULLMANN, C.; KREIL, G. & FRISCHAUF, A.-M., 1983. Nucleotide sequence of cloned cDNA coding for honeybee prepromelittin. *Eur. J. Biochem.*, 135 :123-126.
- ZASLOFF, M., 1987. Magainins, a class of antimicrobial peptides from *Xenopus* skin: Isolation, characterization of two active forms, and partial cDNA sequence of a precursor. *Proc. Natl. Acad. Sci. U.S.A.*, 84 :5449-5453.